Linear stability of a vectorial kinetic relaxation scheme with a central velocity

Clémentine Courtès, Emmanuel Franck

Linear stability of a vectorial kinetic relaxation scheme with a central velocity

Clémentine Courtès[1] and Emmanuel Franck[2]

December 3rd, 2018

Abstract

This article deals with the linear stability of an implicit vectorial kinetic relaxation scheme with a central velocity used to solve numerically some multi-scale hyperbolic systems.
Keywords : linear stability, von Neumann analysis, kinetic relaxation, semi-Lagrangian method, implicit splitting scheme, eigenvalues.

1 Introduction

Hyperbolic systems are often used to model complex physical phenomena such as multi-scale problems. In such problems, characteristic waves do not propagate with the same speed: fast waves interact with slower waves.
Discretizing such physical phenomena is still an open issue. Explicit methods are prohibited due to their very restrictive CFL condition imposed by fastest scales and implicit methods are computational time-consuming and memory cost-consuming due to the inversion of ill-conditioned nonlinear systems. In order to better grasp numerically these multi-scale problems, an alternative is to use kinetic relaxation methods.
The key idea of these kinetic relaxation methods is to consider the unknown of the hyperbolic system as the macroscopic moment of a kinetic distribution function. The main advantage is that the distribution function satisfies a mesoscopic kinetic equation, which is easier to process because it is composed of an advection equation (at constant speeds) combined with a relaxation term, often chosen of Bhatnagar-Gross-Krook type (in short BGK) [2]. The relaxation term enables kinetic equation to tend toward hyperbolic system for an asymptotically small relaxation parameter.
An important degree of freedom in the kinetic relaxation methods is the choice of the number and the values of the constant advection speeds for the distribution function. We follow here the vectorial kinetic relaxation method, introduced in [7, 1], which consists of fixing the same (small) set of advection speeds for each component of the unknown of the hyperbolic system. A suitable choice for multi-scale problems is the one introduced in [4] and mainly developed in [5], where three advection speeds $\lambda_-, \lambda_0, \lambda_+$ are associated to each of the components of the unknown of the hyperbolic system. The central speed λ_0 is added to treat the slowest scale of the physical phenomenon.
From numerical point of view, vectorial kinetic relaxation models are often discretized with numerical schemes which split the advection part from the relaxation term. There seems to be widespread agreement that the relaxation term is treated numerically as a source term. The noticeable difference

[1]Institut de Mathématiques de Toulouse, UMR CNRS 5219, Université de Toulouse, INSA, F-31077 Toulouse, France, clementine.courtes@math.univ-toulouse.fr
[2]INRIA Nancy-Grand Est, équipe TONUS - TOkamaks and NUmerical Simulations and IRMA, UMR CNRS 7501, Université de Strasbourg, France, emmanuel.franck@inria.fr

between numerical schemes mainly comes from the numerical processing of the advection part. It may be treated, for example, by an Exact discrete Transport, as in [6] or by a Semi-Lagrangian method as in [5], which has the advantage to avoid matrices storage and CFL condition. The properties of such a numerical scheme are detailed in [5], particularly for the consistency. The stability analysis is much more difficult and is rather sketchy.
The aim of this current paper is precisely to review all results on that stability property. For simplicity, we restrict our study to the notion of linear stability (or L^2-stability). Note that other notions of stability such as entropic one is briefly discussed in [5]. The outline of the current paper is constructed as follow. Section 2 gathers the notations of the splitting scheme associated to the vectorial kinetic relaxation model with a central velocity. Section 3 is a brief reminder of the notion of L^2-stability. This linear stability issue is raised in Section 4 for a Semi-Lagrangian method for the advection part and in Section 5 for an Exact discrete Transport method.

2 The vectorial kinetic relaxation scheme

Let us consider a 1D linear hyperbolic system $\partial_t \boldsymbol{U} + \partial_x \boldsymbol{F}(\boldsymbol{U}) = 0$, with $\boldsymbol{U}(t,x) \in \mathbb{R}^N$. The flux $\boldsymbol{F}$ is assumed to be linear : $\boldsymbol{F}(\boldsymbol{U}) = A\boldsymbol{U}$ with the square matrix $A \in \mathcal{M}_N(\mathbb{R})$.
The kinetic relaxation representation. By following the notations introduced in [5], a fixed set of velocities $\{\lambda_-, \lambda_0, \lambda_+\}$ with $\lambda_- < \lambda_0 < \lambda_+$ is associated to each of the N components of $\boldsymbol{U}$. Then, $\boldsymbol{U}$ is considered as a macroscopic moment of a kinetic distribution function $\boldsymbol{f} \in \mathbb{R}^{3N}$, which satisfies the following kinetic relaxation equation

$$\partial_t \boldsymbol{f} + \Lambda \partial_x \boldsymbol{f} = \frac{1}{\varepsilon}\left(\boldsymbol{f}^{eq}(\boldsymbol{U}) - \boldsymbol{f}\right). \tag{1}$$

According to the choice of the advection speeds set, we decompose $\boldsymbol{f}$ such as $\boldsymbol{f} = \left(\boldsymbol{f}_-, \boldsymbol{f}_0, \boldsymbol{f}_+\right)^t$, with $\boldsymbol{f}_j = (f_{j,k})_{k\in\{1,\dots,N\}} \in \mathbb{R}^N$ for $j \in \{-,0,+\}$. The left hand side of (1) consists on the advection part with the diagonal matrix $\Lambda = \mathrm{diag}\left(\lambda_-\mathrm{Id}, \lambda_0\mathrm{Id}, \lambda_+\mathrm{Id}\right)$, which contains all the advection speeds (Id is the N-identity matrix). The right hand side of (1) consists on the BGK relaxation part with $\varepsilon > 0$ the relaxation parameter and $\boldsymbol{f}^{eq} = (\boldsymbol{f}_-^{eq}, \boldsymbol{f}_0^{eq}, \boldsymbol{f}_+^{eq})^t$ the equilibrium vector, which is a function of $\boldsymbol{U}$ and which satisfies some consistency properties.
In order to determine $\boldsymbol{f}_-^{eq}$, $\boldsymbol{f}_0^{eq}$ and $\boldsymbol{f}_+^{eq}$, we perform the decentered flux vector splitting detailed in [5]. It consists to decompose the hyperbolic flux $\boldsymbol{F}$ into three parts, which commute each other: $\boldsymbol{F}(\boldsymbol{U}) = \boldsymbol{F}_0^-(\boldsymbol{U}) + \boldsymbol{F}_0^+(\boldsymbol{U}) + \lambda_0\boldsymbol{U}$. In the linear case, the hyperbolic flux $\boldsymbol{F}$ writes $\boldsymbol{F}(\boldsymbol{U}) = A\boldsymbol{U}$ with A a diagonalizable square matrix (because $\boldsymbol{F}$ is hyperbolic) and the previous decomposition is also linear: there exist two commuting diagonalizable square matrices $A_0^\pm$ such that

$$A\boldsymbol{U} = A_0^-\boldsymbol{U} + A_0^+\boldsymbol{U} + \lambda_0\boldsymbol{U}. \tag{2}$$

Decomposition (2) together with $\boldsymbol{U} = \sum_{j\in\{-,0,+\}} \boldsymbol{f}_j$ (since $\boldsymbol{U}$ is the macroscopic moment of $\boldsymbol{f}$) enable to define each $\boldsymbol{f}_j^{eq}$ for $j \in \{-,0,+\}$ as follow, where Id is the identity matrix, (for more details, see [5])

$$\begin{cases} \boldsymbol{f}_-^{eq}(\boldsymbol{U}) = -\dfrac{1}{\lambda_0-\lambda_-}A_0^-\left(\boldsymbol{f}_- + \boldsymbol{f}_0 + \boldsymbol{f}_+\right), \\ \boldsymbol{f}_0^{eq}(\boldsymbol{U}) = \left[\mathrm{Id} - \left(\dfrac{1}{\lambda_+-\lambda_0}A_0^+ - \dfrac{1}{\lambda_0-\lambda_-}A_0^-\right)\right]\left(\boldsymbol{f}_- + \boldsymbol{f}_0 + \boldsymbol{f}_+\right), \\ \boldsymbol{f}_+^{eq}(\boldsymbol{U}) = \dfrac{1}{\lambda_+-\lambda_0}A_0^+\left(\boldsymbol{f}_- + \boldsymbol{f}_0 + \boldsymbol{f}_+\right). \end{cases} \tag{3}$$

The numerical scheme. As in [5], we fix $\Delta t > 0$ and $\Delta x > 0$ the time and space steps and denote $\boldsymbol{f}^n = \left(f^n_{j,k}\right)_{j\in\{-,0,+\},k\in\{1,..,N\}}$ the distribution vector at time $t^n = n\Delta t \in [0,T]$. The numerical scheme chosen to discretize (1) is the following splitting scheme: $\boldsymbol{f}^{n+1} = \mathcal{R}_\varepsilon(\Delta t, \Delta x, \theta) \circ \mathcal{T}(\Delta t, \Delta x)\boldsymbol{f}^n$.
For convenient, we denote $\boldsymbol{f}^* = \mathcal{T}(\Delta t, \Delta x)\boldsymbol{f}^n$.
• The transport step, named $\mathcal{T}(\Delta t, \Delta x)$, may be either a Semi-Lagrangian scheme (SL hereafter), defined by

$$f^*_{j,k}(x) = \mathrm{I}_{\Delta x}\left(f^n_{j,k}\right)(x - \lambda_j \Delta t), \quad \forall j \in \{-,0,+\} \text{ and } \forall k \in \{1,..,N\}, \tag{4}$$

where, for any $g : \mathbb{R} \mapsto \mathbb{R}$, $\mathrm{I}_{\Delta x}(g)$ is a piecewise polynomial interpolation of the values taken by g on the mesh points, or an Exact discrete Transport scheme (ET hereafter), defined by $\mathrm{I}_{\Delta x} = \mathrm{Id}$ (the identity map)

$$f^*_{j,k}(x) = f^n_{j,k}(x - \lambda_j \Delta t), \quad \forall j \in \{-,0,+\} \text{ and } \forall k \in \{1,..,N\}. \tag{5}$$

Remark 1. Assuming an Exact discrete Transport scheme, as Relation (5), leads to a CFL condition since it makes $\frac{\lambda_j \Delta t}{\Delta x}$ be an integer, for all $j \in \{-,0,+\}$.

• The relaxation step, named $\mathcal{R}_\varepsilon(\Delta t, \Delta x, \theta)$, consists on a θ-scheme, with $\theta \in [\frac{1}{2}, 1]$, defined by $\frac{\boldsymbol{f}^{n+1} - \boldsymbol{f}^*}{\Delta t} = \theta \frac{\boldsymbol{f}^{eq}(\boldsymbol{U}^{n+1}) - \boldsymbol{f}^{n+1}}{\varepsilon} + (1-\theta)\frac{\boldsymbol{f}^{eq}(\boldsymbol{U}^*) - \boldsymbol{f}^*}{\varepsilon}$.
Since $\boldsymbol{U}^{n+1} = \boldsymbol{U}^*$ during the relaxation step, cf [5], it may be rewritten in the form:

$$\boldsymbol{f}^{n+1} = \boldsymbol{f}^* + \omega\left(\boldsymbol{f}^{eq}(\boldsymbol{U}^*) - \boldsymbol{f}^*\right) \quad \text{with} \quad \omega = \frac{\Delta t}{\varepsilon + \theta \Delta t} \in [0,2]. \tag{6}$$

The final numerical scheme is thus obtained by combining (4)-(6) for the Semi-Lagrangian choice (or (5)-(6) for the Exact discrete Transport choice).

3 Linear stability

We restrict our study to a linear (or L^2) stability, by a von Neumann analysis.

3.1 A review of L^2-stability

Let G be the amplification matrix of a one-step linear scheme (S) : $\boldsymbol{f}^{n+1} = G(\Delta t, \Delta x)\boldsymbol{f}^n$. We recall the notion of L^2-stability in the following definition.

Definition 3.1. The scheme (S) is L^2-stable if there exists a constant $K > 0$ such that, for all Δt and Δx small enough (and possibly satisfying a CFL condition), for all $n \geq 0$ such that $n\Delta t \leq T$, one has $||\boldsymbol{f}^{n+1}||_{\ell^2} \leq (1 + K\Delta t)||\boldsymbol{f}^n||_{\ell^2}$.

In terms of amplification matrix, the L^2-stability notion translates into the following necessary and sufficient condition : $\sqrt{\rho\left([G(\Delta t, \Delta x)]^* G(\Delta t, \Delta x)\right)} \leq 1 + K\Delta t$, with, for a square matrix G, $\rho(G)$ the spectral radius of G and G^* the adjoint matrix of G. This necessary and sufficient condition is not always easy to verify so we focus only to the sufficient condition of the following proposition.

Proposition 1. ***Sufficient condition :*** *In space Fourier variables $\widehat{\boldsymbol{f}^n}(\xi)$ with $\xi \in [0, \frac{2\pi}{\Delta x}]$, a sufficient condition to ensure the L^2-stability is as follow:*
$\sup_{\xi\in[0,\frac{2\pi}{\Delta x}]} \rho\left(G(\Delta t, \xi)\right) < 1$, *or* $\sup_{\xi\in[0,\frac{2\pi}{\Delta x}]} \rho\left(G(\Delta t, \xi)\right) = 1$ *and the eigenvalues of $G(\Delta t, \xi)$ with modulus equal to 1 are simple.*

3.2 Amplification matrix for the vectorial kinetic relaxation scheme

To deal with the L^2-stability, we have to first compute the amplification matrix.

Reformulation of the relaxation step in the linear case. Since A_0^- and A_0^+ commute and are both diagonalizable, they are diagonalizable in the same basis to obtain $A_0^\pm = B_0 D_0^\pm B_0^{-1}$ with $D_0^\pm$ the diagonal matrices $D_0^\pm = \mathrm{diag}\left(\lambda_k(A_0^\pm)\right)_{k\in\{1,..,N\}}$ and B_0 an invertible matrix. The diagonal term $\lambda_k(A_0^\pm)$ corresponds to the k^{th} eigenvalue of $A_0^\pm$.

Relaxation step (6) is thus rewritten under the following bloc matrices form:

$$\begin{pmatrix} \boldsymbol{f}_-^{n+1} \\ \boldsymbol{f}_0^{n+1} \\ \boldsymbol{f}_+^{n+1} \end{pmatrix} = \mathcal{B}_0 R_\varepsilon(\Delta t,\omega)\mathcal{B}_0^{-1} \begin{pmatrix} \boldsymbol{f}_-^{*} \\ \boldsymbol{f}_0^{*} \\ \boldsymbol{f}_+^{*} \end{pmatrix} \tag{7}$$

with $\mathcal{B}_0 = \mathrm{diag}(B_0, B_0, B_0)$, $\mathcal{B}_0^{-1} = \mathrm{diag}(B_0^{-1}, B_0^{-1}, B_0^{-1})$ and $R_\varepsilon(\Delta t,\omega)$ the relaxation amplification matrix defined by blocs by

$$R_\varepsilon(\Delta t,\omega) = \mathrm{diag}(\mathrm{Id},\mathrm{Id},\mathrm{Id}) + \omega \tilde{R}_\varepsilon(\Delta t),$$

with $\tilde{R}_\varepsilon(\Delta t) =$

$$\begin{pmatrix} -\frac{1}{\lambda_0-\lambda_-}D_0^- - \mathrm{Id} & -\frac{1}{\lambda_0-\lambda_-}D_0^- & -\frac{1}{\lambda_0-\lambda_-}D_0^- \\ \mathrm{Id} - \left(\frac{1}{\lambda_+-\lambda_0}D_0^+ - \frac{1}{\lambda_0-\lambda_-}D_0^-\right) & -\frac{1}{\lambda_+-\lambda_0}D_0^+ + \frac{1}{\lambda_0-\lambda_-}D_0^- & \mathrm{Id} - \left(\frac{1}{\lambda_+-\lambda_0}D_0^+ - \frac{1}{\lambda_0-\lambda_-}D_0^-\right) \\ \frac{1}{\lambda_+-\lambda_0}D_0^+ & \frac{1}{\lambda_+-\lambda_0}D_0^+ & \frac{1}{\lambda_+-\lambda_0}D_0^+ - \mathrm{Id} \end{pmatrix}.$$

Fourier analysis. We introduce the Fourier variable (in space) $\widehat{\boldsymbol{f}^n}$ in $L^2([0, \frac{2\pi}{\Delta x}])$ defined by $\widehat{\boldsymbol{f}^n}(\xi) = \sum_{x\in\{\text{mesh points}\}} \boldsymbol{f}^n(x)e^{ix\xi}$, $\forall \xi \in [0, \frac{2\pi}{\Delta x}]$.

With this Fourier decomposition, transport step of the scheme rewrites :

$$\begin{pmatrix} \widehat{\boldsymbol{f}_-^{*}}(\xi) \\ \widehat{\boldsymbol{f}_0^{*}}(\xi) \\ \widehat{\boldsymbol{f}_+^{*}}(\xi) \end{pmatrix} = \begin{pmatrix} T_-(\Delta t,\xi) & & 0 \\ & T_0(\Delta t,\xi) & \\ 0 & & T_+(\Delta t,\xi) \end{pmatrix} \begin{pmatrix} \widehat{\boldsymbol{f}_-^{n}}(\xi) \\ \widehat{\boldsymbol{f}_0^{n}}(\xi) \\ \widehat{\boldsymbol{f}_+^{n}}(\xi) \end{pmatrix},$$

with

- $T_j(\Delta t,\xi) = T_j^{SL}(\Delta t,\xi)$ the amplification factor of the Semi-Lagrangian scheme (4),
- or $T_j(\Delta t,\xi) = e^{i\lambda_j \Delta t\xi}\mathrm{Id}$ in the case of an Exact discrete Transport scheme (5).

Since relaxation step does not depend on the space, Relation (7) is also true with $\widehat{\boldsymbol{f}^{n+1}}$ (resp. $\widehat{\boldsymbol{f}^{*}}$) instead of $\boldsymbol{f}^{n+1}$ (resp. $\boldsymbol{f}^{*}$).

Eventually, the total amplification matrix is equal to

$$G(\Delta t,\xi,\varepsilon,\omega) = \mathcal{B}_0 R_\varepsilon(\Delta t,\omega)\mathcal{B}_0^{-1} \begin{pmatrix} T_-(\Delta t,\xi) & & 0 \\ & T_0(\Delta t,\xi) & \\ 0 & & T_+(\Delta t,\xi) \end{pmatrix}, \tag{8}$$

with $T_j(\Delta t,\xi) = T_j^{SL}(\Delta t,\xi)$ for the scheme (4)-(6), or with $T_j(\Delta t,\xi) = e^{i\lambda_j\Delta t\xi}\mathrm{Id}$ for the scheme (5)-(6), for $j \in \{-,0,+\}$.

4 Linear stability for the Semi-Lagrangian step

Let us ensure our first linear stability result.

Proposition 2. *Let the hyperbolic flux* $\boldsymbol{F}$ *be linear and be decomposed into* $\boldsymbol{F}(\boldsymbol{U}) = A\boldsymbol{U} = A_0^-\boldsymbol{U} + A_0^+\boldsymbol{U} + \lambda_0\boldsymbol{U}$, *with* A_0^+ *and* A_0^- *two commuting and diagonalizable matrices.*
The numerical scheme (4)-(6) *with the advection speeds set* $\{\lambda_-, \lambda_0, \lambda_+\}^N$, *with* $\lambda_- < \lambda_0 < \lambda_+$ *and equilibrium* (3) *is* L^2*-stable on the following conditions :*

- $\omega \in [0, 1]$,
- $\mathrm{Id} - \left(\frac{1}{\lambda_+ - \lambda_0} A_0^+ - \frac{1}{\lambda_0 - \lambda_-} A_0^-\right)$ *is a positive semidefinite matrix,*
- A_0^+ *(resp.* A_0^-*) is a positive (resp. negative) semidefinite matrix.*

Remark 2. Note that here a matrix is said to be positive semidefinite (resp. negative semidefinite) if all its eigenvalues are nonnegative (resp. nonpositive).

Remark 3. Sufficient conditions which involve in Proposition 2 are exactly the same as the ones required for the entropy stability property, proved in [5].

To prove Proposition 2, we follow the main guidelines of [6] which suggest to study the Gershgorin discs of the amplification matrix, for more details see [9].

Definition 4.1. Let $G = (g_{ij})_{i,j} \in \mathcal{M}_N(\mathbb{C})$ be a complex square matrix. The k^{th}-Gershgorin disc corresponds to the disc $\mathcal{D}_k = \{z \in \mathbb{C}, |g_{kk} - z| \leq \sum_{j \neq k} |g_{jk}|\}$, for $k \in \{1, ..., N\}$.

Theorem 4.2 (Gershgorin's theorem)**.** *Let* $G = (g_{ij})_{i,j} \in \mathcal{M}_N(\mathbb{C})$ *be a complex square matrix. Every eigenvalue of* G *belongs at least to one Gershgorin disc of* G.

Proof of Proposition 2. As the Semi-Lagrangian step is unconditionally stable [3], we may omit the Semi-Lagrangian amplification factors $T_j^{SL}(\Delta t, \xi)$ for $j \in \{-, 0, +\}$ in our study. It only remains to consider eigenvalues of $R_\varepsilon(\Delta t, \omega)$.

Let λ be an eigenvalue of $R_\varepsilon(\Delta t, \omega) = (r_{\varepsilon,ij})_{i,j \in \{1,...,3N\}}$. According to Theorem 4.2, there exists $\bar{k} \in \{1, ..., 3N\}$ such that $|r_{\varepsilon,\bar{k}\bar{k}} - \lambda| \leq \sum_{j \neq \bar{k}} |r_{\varepsilon,j\bar{k}}|$. Then by a triangular inequality, $|\lambda| \leq |r_{\varepsilon,\bar{k}\bar{k}} - \lambda| + |r_{\varepsilon,\bar{k}\bar{k}}| \leq \sum_{j=1}^{3N} |r_{\varepsilon,j\bar{k}}|$. However, one has

- for $\bar{k} \in \{1, ..., N\}$

$$\sum_{j=1}^{3N} |r_{\varepsilon,j\bar{k}}| = \left|1 - \omega\frac{\lambda_{\bar{k}}(A_0^-)}{\lambda_0 - \lambda_-} - \omega\right| + \left|\omega - \omega\left(\frac{\lambda_{\bar{k}}(A_0^+)}{\lambda_+ - \lambda_0} - \frac{\lambda_{\bar{k}}(A_0^-)}{\lambda_0 - \lambda_-}\right)\right| + \left|\omega\frac{\lambda_{\bar{k}}(A_0^+)}{\lambda_+ - \lambda_0}\right|,$$

- for $\bar{k} \in \{N + 1, ..., 2N\}$

$$\sum_{j=1}^{3N} |r_{\varepsilon,j\bar{k}}| = \left|-\omega\frac{\lambda_{\bar{k}}(A_0^-)}{\lambda_0 - \lambda_-}\right| + \left|1 - \omega + \omega\left(1 - \left(\frac{\lambda_{\bar{k}}(A_0^+)}{\lambda_+ - \lambda_0} - \frac{\lambda_{\bar{k}}(A_0^-)}{\lambda_0 - \lambda_-}\right)\right)\right| + \left|\omega\frac{\lambda_{\bar{k}}(A_0^+)}{\lambda_+ - \lambda_0}\right|,$$

- for $\bar{k} \in \{2N + 1, ..., 3N\}$

$$\sum_{j=1}^{3N} |r_{\varepsilon,j\bar{k}}| = \left|-\omega\frac{\lambda_{\bar{k}}(A_0^-)}{\lambda_0 - \lambda_-}\right| + \left|\omega - \omega\left(\frac{\lambda_{\bar{k}}(A_0^+)}{\lambda_+ - \lambda_0} - \frac{\lambda_{\bar{k}}(A_0^-)}{\lambda_0 - \lambda_-}\right)\right| + \left|1 + \omega\frac{\lambda_{\bar{k}}(A_0^+)}{\lambda_+ - \lambda_0} - \omega\right|.$$

Hypotheses of Proposition 2 enable to remove modulus in the previous relations and to obtain for all $\bar{k} \in \{1, ..., 3N\}$, $\sum_{j=1}^{3N} |r_{\varepsilon,j\bar{k}}| = 1$.

Hence, all eigenvalues of $R_\varepsilon(\Delta t, \omega)$ have a modulus smaller than 1 which implies the L^2- stability of the numerical scheme (4)-(6).

□

Example 1. Here is a non-exhaustive list of flux decompositions which enable to obtain a L^2-stable scheme for scalar case when $F(u) = au = a_0^- u + a_0^+ u + \lambda_0 u$:

• (Rusanov) We choose $\lambda_- \leq \min(0, a)$, $\lambda_0 = 0$ and $\max(0, a) \leq \lambda_+$ and we define $a_0^+ = \lambda_+ \left(\frac{a-\lambda_-}{\lambda_+-\lambda_-}\right)$ and $a_0^- = -\lambda_- \left(\frac{a-\lambda_+}{\lambda_+-\lambda_-}\right)$. A particular Rusanov decomposition consists to choose $-\lambda_- = \lambda_+ = |a|$.

• (Upwind) We choose $\lambda_- \leq \min(\lambda_0, a)$ and $\max(\lambda_0, a) \leq \lambda_+$ and we define $a_0^+ = \mathbb{1}_{a>\lambda_0}(a - \lambda_0)$ and $a_0^- = \mathbb{1}_{a<\lambda_0}(a - \lambda_0)$, with $\mathbb{1}$ the indicator function.

• (Lax-Wendroff) We choose $-\lambda_- = \lambda_+ = \lambda > 0$ with $\sqrt{\alpha}|a| \leq \lambda \leq \alpha|a|$ and $\lambda_0 = 0$ and we define $a_0^\pm = \frac{1}{2}\left(a \pm \alpha\frac{a^2}{\lambda}\right)$ with $\alpha \in [1, 2]$.

For more details about these flux decompositions, we refer the readers to [5].

Remark 4. Proposition 2 is also valid for the scheme (5)-(6) (with an Exact discrete Transport step). However, the following section improves those results.

5 Linear stability for the Exact discrete Transport step

For simplicity, we focus only on the scalar linear case: $\partial_t u + a\partial_x u = 0$ with $u(t, x) \in \mathbb{R}$ and $a \in \mathbb{R}$, which implies $\mathcal{B}_0 = 1$ in (8). The Exact discrete Transport step enables to improve sufficient conditions of Proposition 2, in particular in the range of admissible ω.

Proposition 3. *Let the scalar hyperbolic flux F be linear and be decomposed into $F(u) = au = a_0^- u + a_0^+ u + \lambda_0 u$.*
The numerical scheme (5)-(6) with the advection speeds set $\{\lambda_-, \lambda_0, \lambda_+\}$, with $\lambda_- < \lambda_0 < \lambda_+$ and equilibrium (3) is L^2-stable on the following conditions :

- $\omega \in [0, 2]$,
- $1 - \left(\frac{a_0^+}{\lambda_+-\lambda_0} - \frac{a_0^-}{\lambda_0-\lambda_-}\right) \geq 0$,
- $a_0^+ \geq 0$ *and* $a_0^- \leq 0$,
- *One of the three following equalities is satisfied :* $a_0^+ = 0$ *or* $a_0^- = 0$ *or* $1 - \left(\frac{a_0^+}{\lambda_+-\lambda_0} - \frac{a_0^-}{\lambda_0-\lambda_-}\right) = 0$.

Remark 5. As the scalar case is considered here, condition $1 - \left(\frac{a_0^+}{\lambda_+-\lambda_0} - \frac{a_0^-}{\lambda_0-\lambda_-}\right) \geq 0$ may be written in the simplest form: $\lambda_- \leq a \leq \lambda_+$. In deed, in the scalar case,

$$1 - \left(\frac{a_0^+}{\lambda_+ - \lambda_0} - \frac{a_0^-}{\lambda_0 - \lambda_-}\right) = \frac{\lambda_+ - a}{\lambda_+ - \lambda_0} + \frac{a_0^-(\lambda_+ - \lambda_-)}{(\lambda_+ - \lambda_0)(\lambda_0 - \lambda_-)} = \frac{a - \lambda_-}{\lambda_0 - \lambda_-} - \frac{a_0^+(\lambda_+ - \lambda_-)}{(\lambda_+ - \lambda_0)(\lambda_0 - \lambda_-)}.$$

The nonnegativity of this quantity together with the hypotheses $\lambda_- < \lambda_0 < \lambda_+$, $a_0^+ \geq 0$ and $a_0^- \leq 0$ implie that $\lambda_- \leq a \leq \lambda_+$.

Instead of using Gershgorin discs to prove Proposition 3, we need a more specific tool to localize the eigenvalues and we use Rouché's theorem, as suggested in [8].

Theorem 5.1 (Rouché's theorem)**.** *Let γ be a closed simple path in $\Omega \subset \mathbb{C}$ and assume that γ has an interior. Let f and g be holomorphic (analytic) on Ω and $|f(\zeta) - g(\zeta)| < |f(\zeta)|$ for all ζ on γ. Then f and g have the same number of zeros in the interior of γ.*

Proof of Proposition 3. Let us define $\mathcal{A}_0^+ = \frac{a_0^+}{\lambda_+ - \lambda_0}$ and $\mathcal{A}_0^- = \frac{a_0^-}{\lambda_0 - \lambda_-}$. In the scalar case, the amplification matrix writes $G(\Delta t, \xi, \varepsilon, \omega) =$

$$\begin{pmatrix} \left(1 - \omega\mathcal{A}_0^- - \omega\right) e^{i\lambda_- \Delta t\xi} & -\omega\mathcal{A}_0^- e^{i\lambda_0 \Delta t\xi} & -\omega\mathcal{A}_0^- e^{i\lambda_+ \Delta t\xi} \\ \omega\left(1 - \left(\mathcal{A}_0^+ - \mathcal{A}_0^-\right)\right) e^{i\lambda_- \Delta t\xi} & \left(1 - \omega\left(\mathcal{A}_0^+ - \mathcal{A}_0^-\right)\right) e^{i\lambda_0 \Delta t\xi} & \omega\left(1 - \left(\mathcal{A}_0^+ - \mathcal{A}_0^-\right)\right) e^{i\lambda_+ \Delta t\xi} \\ \omega\mathcal{A}_0^+ e^{i\lambda_- \Delta t\xi} & \omega\mathcal{A}_0^+ e^{i\lambda_0 \Delta t\xi} & \left(1 + \omega\mathcal{A}_0^+ - \omega\right) e^{i\lambda_+ \Delta t\xi} \end{pmatrix}.$$

The characteristic polynomial of G is as follow: $\chi(X) = \mu_0 + \mu_1 X + \mu_2 X^2 - X^3$, with

$$\begin{aligned} \mu_0 &= (1-\omega)^2 e^{i(\lambda_+ + \lambda_0 + \lambda_-)\Delta t\xi}, \\ \mu_1 &= (1-\omega)\Big[(-1 - \omega\mathcal{A}_0^-) e^{i(\lambda_0 + \lambda_+)\Delta t\xi} + (-1 + \omega\mathcal{A}_0^+) e^{i(\lambda_0 + \lambda_-)\Delta t\xi} \\ &\qquad + (-1 + \omega(1 - \mathcal{A}_0^+ + \mathcal{A}_0^-)) e^{i(\lambda_+ + \lambda_-)\Delta t\xi}\Big], \\ \mu_2 &= (1 - \omega + \omega\mathcal{A}_0^+) e^{i\lambda_+ \Delta t\xi} + (1 - \omega + \omega(1 - \mathcal{A}_0^+ + \mathcal{A}_0^-)) e^{i\lambda_0 \Delta t\xi} \\ &\qquad + (1 - \omega - \omega\mathcal{A}_0^-) e^{i\lambda_- \Delta t\xi}. \end{aligned}$$

In the three particular studied cases, the previous characteristic polynomial writes

$$\chi(X) = \left\{(1-\omega)e^{i\nu_1 \Delta t\xi} - X\right\} \tilde{\chi}(X), \tag{9}$$

where $\tilde{\chi}(X) = (1-\omega)e^{i(\nu_2+\nu_3)\Delta t\xi} - X e^{i\left(\frac{\nu_2+\nu_3}{2}\right)\Delta t\xi}\left[(2-\omega)\cos(\Xi) + i\omega\eta\sin(\Xi)\right] + X^2$, with

- **Case $a_0^+ = 0$:** $\nu_1 = \lambda_+,\ \nu_2 = \lambda_0,\ \nu_3 = \lambda_-,\ \Xi = \frac{\lambda_0 - \lambda_-}{2}\Delta t\xi,\ \eta = 1 + 2\frac{a_0^-}{\lambda_0 - \lambda_-}$,
- **Case $a_0^- = 0$:** $\nu_1 = \lambda_-,\ \nu_2 = \lambda_0,\ \nu_3 = \lambda_+,\ \Xi = \frac{\lambda_0 - \lambda_+}{2}\Delta t\xi,\ \eta = 1 - 2\frac{a_0^+}{\lambda_+ - \lambda_0}$,
- **Case $1 - \frac{a_0^+}{\lambda_+ - \lambda_0} + \frac{a_0^-}{\lambda_0 - \lambda_-} = 0$:** $\nu_1 = \lambda_0,\ \nu_2 = \lambda_+,\ \nu_3 = \lambda_-,\quad \Xi = \frac{\lambda_+ - \lambda_-}{2}\Delta t\xi,\ \eta = \frac{a_0^+}{\lambda_+ - \lambda_0} + \frac{a_0^-}{\lambda_0 - \lambda_-}$.

Particular cases $\omega = 0$ or $\omega = 2$: In these two cases, the three roots are transparent:
$\left\{e^{i\nu_1 \Delta t\xi}, \left[\cos(\Xi) \pm i\sin(\Xi)\right] e^{i\left(\frac{\nu_2+\nu_3}{2}\right)\Delta t\xi}\right\} = \left\{e^{i\nu_1 \Delta t\xi}, e^{i\nu_2 \Delta t\xi}, e^{i\nu_3 \Delta t\xi}\right\}$ for $\omega = 0$ and
$\left\{-e^{i\nu_1 \Delta t\xi}, \left[\pm\sqrt{1 - \eta^2\sin^2(\Xi)} + i\eta\sin(\Xi)\right] e^{i\left(\frac{\nu_2+\nu_3}{2}\right)\Delta t\xi}\right\}$ for $\omega = 2$.
Equality (2) in the scalar case enables to simplify η: $\eta = -\frac{\nu_2 - 2a + \nu_3}{\nu_2 - \nu_3}$. Conditions $\lambda_- < \lambda_0 < \lambda_+$, $\lambda_- \leq a \leq \lambda_+$, $a_0^- \leq 0$ and $a_0^+ \geq 0$ imply that $\eta \in [-1, 1]$. Thus, the square root $\sqrt{1 - \eta^2\sin^2(\Xi)}$ is well defined. If $-1 < \eta < 1$, the roots are simple since $1 - \eta^2\sin^2(\Xi) \neq 0$. Otherwise, the roots for $\omega = 2$ simplify into $\left\{-e^{i\nu_1 \Delta t\xi}, e^{i\nu_2 \Delta t\xi}, -e^{i\nu_3 \Delta t\xi}\right\}$ if $\eta = 1$ and $\left\{-e^{i\nu_1 \Delta t\xi}, -e^{i\nu_2 \Delta t\xi}, e^{i\nu_3 \Delta t\xi}\right\}$ if $\eta = -1$.

To conclude with $\omega \in \{0, 2\}$, all these roots have a modulus equal to 1 and are simple.

General case $\omega \in]0, 2[$: Obviously, according to (9), one of the three roots of χ is $(1-\omega)e^{i\nu_1 \Delta t\xi}$ which has a modulus strictly smaller than 1 if $\omega \in]0, 2[$. We have to determine the two other roots.

• If $\Xi \equiv 0[\pi]$, $\tilde{\chi}$ writes $\tilde{\chi}_\pm(X) := (1-\omega)e^{i(\nu_2+\nu_3)\Delta t\xi} \pm [2-\omega]e^{i\left(\frac{\nu_2+\nu_3}{2}\right)\Delta t\xi} X + X^2$.

The roots of $\tilde{\chi}_-$ are $\frac{(2-\omega)\pm\omega}{2}e^{i\left(\frac{\nu_2+\nu_3}{2}\right)\Delta t\xi}$ and those of $\tilde{\chi}_+$ are $\frac{-(2-\omega)\pm\omega}{2}e^{i\left(\frac{\nu_2+\nu_3}{2}\right)\Delta t\xi}$. They are all simple and their modulus are equal to 1 or to $|1-\omega|<1$ since $\omega\in]0,2[$.

• If $\Xi\not\equiv 0[\pi]$, we define ψ such as $\psi(X):=\tilde{\chi}(X)$ for $\omega=1$, which gives

$$\psi(X)=X\left[-e^{i\left(\frac{\nu_2+\nu_3}{2}\right)\Delta t\xi}\left[\cos(\Xi)+i\eta\sin(\Xi)\right]+X\right].$$

The key point is to compare the zeros of both $\tilde{\chi}$ and ψ in the unit ball, as in [8].
Roots of ψ. The roots of ψ are $X_1=0$ and $X_2=e^{i\left(\frac{\nu_2+\nu_3}{2}\right)\Delta t\xi}\left[\cos(\Xi)+i\eta\sin(\Xi)\right]$.
The modulus of X_2 is $|X_2|^2=\cos^2(\Xi)\left[1-\eta^2\right]+\eta^2\in\left[\eta^2,1\right[$, since $\eta\in[-1,1]$.
Since $\Xi\not\equiv 0[\pi]$, $|X_2|$ does not be equal to 1. The function ψ has thus two roots strictly contained in the open unit ball.
Comparison between $\tilde{\chi}$ and ψ on the unit circle. For $\theta\in\mathbb{R}$, one has

$$|\tilde{\chi}(e^{i\theta})-\psi(e^{i\theta})|=|1-\omega|\left|e^{i(\nu_2+\nu_3)\Delta t\xi}-e^{i(\theta+\left(\frac{\nu_2+\nu_3}{2}\right)\Delta t\xi)}\left[\cos(\Xi)-i\eta\sin(\Xi)\right]\right|.$$

Multiplying by $|-e^{-i\left(\theta+\left(\frac{\nu_2+\nu_3}{2}\right)\Delta t\xi\right)}|$ and taking the complex conjugate give

$$|\tilde{\chi}(e^{i\theta})-\psi(e^{i\theta})|=|1-\omega|\left|-e^{-i\left(\frac{\nu_2+\nu_3}{2}\right)\Delta t\xi+i\theta}+\left[\cos(\Xi)+i\eta\sin(\Xi)\right]\right|.$$

Computation of ψ on the unit circle. One has $|\psi(e^{i\theta})|=$
$\left|-e^{i\left(\frac{\nu_2+\nu_3}{2}\right)\Delta t\xi}\left[\cos(\Xi)+i\eta\sin(\Xi)\right]+e^{i\theta}\right|=\left|\cos(\Xi)+i\eta\sin(\Xi)-e^{-i\left(\frac{\nu_2+\nu_3}{2}\right)\Delta t\xi+i\theta}\right|$. The latest equality is obtained by a multiplication by $|-e^{-i\left(\frac{\nu_2+\nu_3}{2}\right)\Delta t\xi}|$.
Use of Rouché's theorem 5.1. One chooses the closed simple path γ be equal to the unit circle. Since $\omega\in]0,2[$, one has $|\tilde{\chi}(e^{i\theta})-\psi(e^{i\theta})|=|1-\omega||\psi(e^{i\theta})|<|\psi(e^{i\theta})|$ for all $\theta\in\mathbb{R}$. By Rouché's theorem 5.1, $\tilde{\chi}$ has the same number of roots in the open unit ball than ψ.

All in all, each case of ω leads to three roots of χ with modulus strictly less than 1 or equal to 1 and simple. The L^2-stability is thus a consequence of Proposition 1.

□

Example 2. The three first flux decompositions of Example 1 satisfy hypotheses of Proposition 3. They are also satisfied by the Lax-Wendroff decomposition only with the extremal choice $|a|=\lambda/\alpha$ or $|a|=\lambda/\sqrt{\alpha}$. (Note that the L^2-stability may be proved by a directe computation with $\alpha=1$ and $\lambda\geq|a|$ in the particular choice $\omega=1$).

References

[1] D. Aregba-Driollet and R. Natalini, Discrete kinetic schemes for systems of conservation laws, in *Hyperbolic Problems: Theory, Numerics, Applications. International Series of Numerical Mathematics* (vol 129), Birkhäuser, Basel, (1999), 1–10.

[2] P. L. Bhatnagar, E. P. Gross and M. Krook, A model for collision processes in gases. I. Small amplitude processes in charged and neutral one-component systems, *Physical review*, **94** (1954), 511–525.

[3] F. Charles, B. Després and M. Mehrenberger, Enhanced convergence estimates for semi-Lagrangian schemes. Application to the Vlasov–Poisson equation, *SIAM Journal on Numerical Analysis*, **51** (2013), 840–863.

[4] D. Coulette, E. Franck, P. Helluy, M. Mehrenberger and L. Navoret, High-order implicit palindromic discontinuous Galerkin method for kinetic-relaxation approximation, preprint, (2018).

[5] D. Coulette, C. Courtès, E. Franck and L. Navoret, Vectorial kinetic relaxation model with central velocity. Application to implicit relaxation schemes, preprint, (2018).

[6] B. Graille, Approximation of mono-dimensional hyperbolic systems: A lattice Boltzmann scheme as a relaxation method, *Journal of Computational Physics*, **266** (2014), 74–88.

[7] R. Natalini, A discrete kinetic approximation of entropy solutions to multidimensional scalar conservation laws, *Journal of Differential Equations*, **148** (1998), 292 – 317.

[8] M. Rheinländer, Stability and multiscale analysis of an advective lattice Boltzmann scheme, *Progress in Computational Fluid Dynamics*, **8** (2008), 56–68.

[9] D. Serre, *Matrices: Theory and Applications*, Graduate Texts in Mathematics, vol. 216, Springer New York, 2010.